



# Three-dimensional descriptors for aminergic GPCRs: dependence on docking conformation and crystal structure

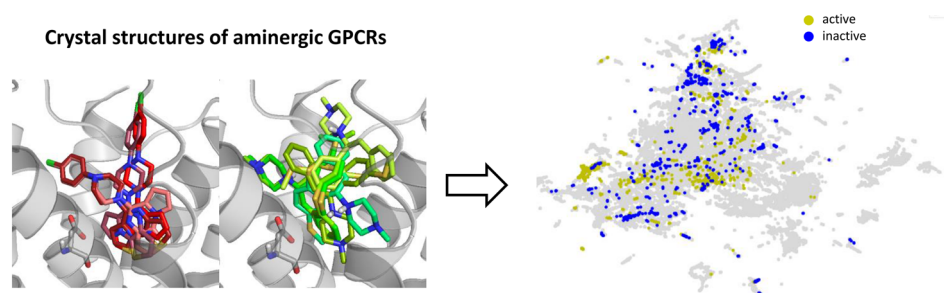
Stanisław Jastrzębski<sup>1</sup> · Igor Sieradzki<sup>1</sup> · Damian Leśniak<sup>1</sup> · Jacek Tabor<sup>1</sup> · Andrzej J. Bojarski<sup>2</sup> · Sabina Podlewska<sup>2</sup> 

Received: 3 October 2018 / Accepted: 12 November 2018 / Published online: 27 November 2018  
© The Author(s) 2018

## Abstract

Three-dimensional descriptors are often used to search for new biologically active compounds, in both ligand- and structure-based approaches, capturing the spatial orientation of molecules. They frequently constitute an input for machine learning-based predictions of compound activity or quantitative structure–activity relationship modeling; however, the distribution of their values and the accuracy of depicting compound orientations might have an impact on the power of the obtained predictive models. In this study, we analyzed the distribution of three-dimensional descriptors calculated for docking poses of active and inactive compounds for all aminergic G protein-coupled receptors with available crystal structures, focusing on the variation in conformations for different receptors and crystals. We demonstrated that the consistency in compound orientation in the binding site is rather not correlated with the affinity itself, but is more influenced by other factors, such as the number of rotatable bonds and crystal structure used for docking studies. The visualizations of the descriptors distributions were prepared and made available online at [http://chem.gmum.net/vischem\\_stability](http://chem.gmum.net/vischem_stability), which enables the investigation of chemical structures referring to particular data points depicted in the figures. Moreover, the performed analysis can assist in choosing crystal structure for docking studies, helping in selection of conditions providing the best discrimination between active and inactive compounds in machine learning-based experiments.

## Graphical abstract



**Keywords** Aminergic GPCRs · Docking · Three-dimensional descriptors · Crystal structure · Machine learning

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11030-018-9894-4>) contains supplementary material, which is available to authorized users.

✉ Sabina Podlewska  
smusz@if-pan.krakow.pl

<sup>1</sup> Faculty of Mathematics and Computer Science, Jagiellonian University, S. Łojasiewicza Street 6, 30-048 Kraków, Poland

<sup>2</sup> Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, Smętna Street 12, 31-343 Kraków, Poland

## Introduction

Computational methods are an indispensable part of the drug design process, supporting the search for new compounds with desired biological activity. One of the most popular in silico strategies is virtual screening (VS), in which large libraries of compounds (commercially available or generated computationally using various combinatorial approaches) undergo nonexperimental evaluation [1–3]. Usually, one of the first filters in VS cascades is connected with the application of *ligand-based* approaches [4], when only structures of compounds with already determined affinity for a particular receptor are used for making predictions for new compounds. The structures that successfully pass to subsequent filtering steps are typically assessed by the confrontation of the features they possess with pharmacophore models constructed for a particular target [5, 6] and/or are docked to the binding site of the respective protein (*structure-based VS* [7]).

One of the fundamental issues connected with the application of various computational methods in the search for new drug candidates is the provision of proper conversion of the chemical structure to a form that can be handled by computer methods. The most popular way to capture chemical information is the application of numerical descriptors or fingerprints [8–11]. The former method usually characterizes the physicochemical properties of compounds, and the examples of such descriptors are as follows: molecular weight, octanol–water partition coefficient ( $\log P$ ),  $pK_a$ , number of hydrogen bond donors, number of hydrogen bond acceptors, number of atoms of a particular type, number of bonds of a particular type, atomic charges, polarity, molecular volume, etc. On the other hand, fingerprints are a result of the translation of compounds into the form of a bit string. Their two main groups can be distinguished: key-based and of a hashed type. Key-based fingerprinting annotates the presence (1) or absence (0) of particular chemical moieties in the molecule, whereas in the hashed fingerprints, bit strings are formed from the molecular graph, on the basis of which paths up to a fixed length are generated (subsequently, starting from each atom), and a hashing function is applied to encapsulate the structural information in the string. Fingerprints are used not only for chemical structure characterization, but they are also applied for description of ligand–receptor complexes obtained in docking. They provide information about the interaction of a ligand with particular amino acids of a protein, as in the case of interaction fingerprints (IFts) [12] and structural interaction fingerprints (SIFts) [13]. The most important advantages of fingerprints are the relative simplicity, low computational costs connected with their generation, and simplicity of making comparisons between two 0–1 strings. The latter procedure can be carried out with the use of various similarity coeffi-

cients or by application of machine learning approaches [14, 15].

The *ligand-based* approaches usually make also use of various molecular descriptors. They are generated from the two-dimensional (2d) structure of the compound (2d descriptors), or they use their spatial orientation obtained either in minimization of a ligand solely or using the docking poses—three-dimensional (3d) descriptors [16, 17].

Predictive models (constructed most often with the use of machine learning methods of various complexities) based on such representations can be of great help in the search for new active compounds. Such approaches are widely known for the ability to deal with a large amount of information in a fast and efficient way, so the application of machine learning is rapidly growing also in the field of computer-aided drug design, mainly due to a significant increase in amount of data also in the field of cheminformatics and pharmacy. However, despite the great power possessed by various algorithms in the identification of new potentially active compounds, their performance strongly depends on experimental conditions, such as training set compositions, compounds representation, and parameters of particular learning algorithms themselves. Therefore, a multifactor optimization needs to be performed in order to obtain the optimal predictive power of such methods [18, 19].

In this study, we focused on one of the above-mentioned parameters influencing the performance of machine learning methods performance, that is, the compounds representation, 3d descriptors in particular. The variations of their values, depending on the input used for their generation (docking pose), were analyzed. It was examined whether the consistency of 3d descriptor values obtained for different orientations in the binding site of the same compound is correlated with the compound activity. All existing crystal structures of aminergic G protein-coupled receptors (GPCRs) were used for docking of known ligands and compounds with confirmed inactivity toward the respective proteins. The 3d descriptor values calculated for poses obtained from docking for active and inactive compounds were compared with those descriptors that were generated from compounds with minimized energies (prepared in LigPrep for docking) and compared with the analysis of variations in atom positions in the docking poses of a particular compound. The results of such study were also made available online ([http://chem.gmunet/vischem\\_stability](http://chem.gmunet/vischem_stability)), together with the possibility to manually analyze chemical structures referring to a particular data point. The differences in the distribution of 3d descriptor values indicate that not all crystal structures are effective in this type of experiments (docking followed by automatic analysis of its results), and the prepared tool can be of great help during the design of such studies, especially by assisting in the selection of crystal structures for docking that would provide optimal performance of machine learning methods (when the

**Table 1** Crystal structures used in the study

Target	Crystal structure	Co-crystallized ligand	Resolution (Å)
5-HT <sub>1B</sub>	4IAQ [24]	Dihydroergotamine	2.8
	4IAR [24]	Ergotamine	2.7
	5V54 [25]	CHEMBL428892	3.9
5-HT <sub>2B</sub>	4IB4 [26]	Ergotamine	2.7
	4NC3 [27]	Ergotamine	2.8
	5TUD [28]	Ergotamine	3.0
	5TVN [29]	Lysergide	2.9
	6BQG [30]	Ergotamine	3.0
5-HT <sub>2C</sub>	6BQH [30]	Ritanserlin	2.7
	5CXV [31]	CHEMBL258622	2.7
ACM <sub>1</sub>	3UON [32]	(R)-(–)-QNB	3.0
ACM <sub>2</sub>	4MQS [33]	Iperoxo	3.5
	4MQT [33]	LY2119620	3.7
		Iperoxo	
ACM <sub>3</sub>	4DAJ [33]	Tiotropium	3.4
	4U14 [34]	Tiotropium	3.6
	4U15 [34]	Tiotropium	2.8
	4U16 [34]	Methscopolamine	3.7
ACM <sub>4</sub>	5DSG [35]	Tiotropium	2.6
Beta1	2VT4 [36]	(S)-Cyanopindolol	2.7
	2Y00 [36]	Dobutamine	2.5
	2Y01 [36]	Dobutamine	2.6
	2Y02 [36]	Carmoterol	2.6
	2Y03 [36]	Levisoprenaline	2.9
	2Y04 [36]	Levalbuterol	3.1
	2YCW [37]	Timolol	3.0
	2YCX [37]	(S)-Cyanopindolol	3.3
	2YCY [37]	(S)-Cyanopindolol	3.2
	2YCZ [37]	Timolol	3.7
	3ZPQ [38]	4-(1-Piperazinyl)-1H-indole	2.8
	3ZPR [38]	4-Methyl-2-piperazin-1-yl-quinoline	2.7
	4AMI [39]	Bucindolol	3.2
	4AMJ [39]	(S)-Carvedilol	2.3
	4BVN [40]	(S)-Cyanopindolol	2.1
	4GPO [41]	–	3.5
	5A8E [42]	7-Methylcyanopindolol	2.4
	5F8U [43]	(S)-Cyanopindolol	3.4
Beta2	2R4R [44]	–	3.4
	2R4S [44]	–	3.4
	2RH1 [45]	Carazolol	2.4
	3D4S [46]	Timolol	2.8
	3KJ6 [47]	–	3.4
	3NY8 [48]	AC1NSK5 N	2.8
	3NY9 [48]	CHEMBL1233771	2.8
	3NYA [48]	(–)-Alprenolol	3.2
	3P0G [49]	BI-167107	3.5
	3PDS [50]	FAUC50	3.5

**Table 1** continued

Target	Crystal structure	Co-crystallized ligand	Resolution (Å)
	4GBR [51]	Timolol	4.0
	5D5A [52]	Carazolol	2.5
	5D5B [52]	Carazolol	3.8
	5D6L [53]	Carazolol	3.2
	5X7D [54]	Carazolol PubchemCID: 129318963	2.7
D <sub>2</sub>	6CM4 [55]	Risperidone	2.9
D <sub>3</sub>	3PBL [56]	Eticlopride	2.9
D <sub>4</sub>	5WIU [57]	Nemonapride	2.0
	5WIV [57]	Nemonapride	2.1
H <sub>1</sub>	3RZE [58]	Doxepin	3.1

**Table 2** Number of compounds from each dataset used in the study

Receptor	Number of compounds before LigPrep		Number of compounds after LigPrep	
	Actives	Inactives	Actives	Inactives
5-HT <sub>1B</sub>	535	435	1174	758
5-HT <sub>2B</sub>	416	311	670	540
5-HT <sub>2C</sub>	1219	1028	1849	1664
ACM <sub>1</sub>	703	964	1335	1000
ACM <sub>2</sub>	727	687	1426	1287
ACM <sub>3</sub>	1054	548	1861	979
ACM <sub>4</sub>	176	352	404	562
Beta1	174	509	264	701
Beta2	230	395	386	556
D <sub>2</sub>	3521	2741	6480	4980
D <sub>3</sub>	2761	799	5203	1272
D <sub>4</sub>	1198	473	1902	827
H <sub>1</sub>	610	587	1301	922

distribution of descriptor values for active and inactive compounds will be too similar, the machine learning algorithms would also face difficulties in making distinction between these two groups of molecules).

## Methods

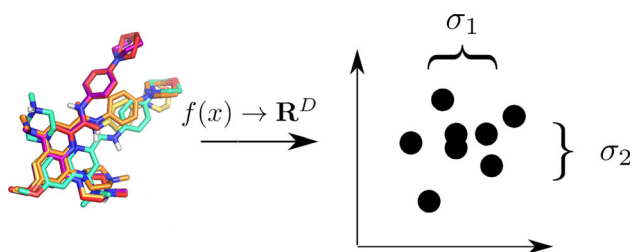
The compounds with experimentally determined activity/inactivity were fetched from the ChEMBL database [20] according to the previously described protocol [21]: Only data produced on human- and rat-cloned receptors were considered, sets of active compounds included structures with  $K_i$  parameter values below 100 nM (also activities

expressed as  $IC_{50}$  were taken into account, assuming that  $K_i = IC_{50}/2$ ) [22], and the sets of inactive compounds included structures with  $K_i$  (or equivalent activity parameter) values above 1000 nM. The considered targets were all aminergic GPCRs for which crystal structures are available, including serotonin receptors 5-HT<sub>1B</sub>, 5-HT<sub>2B</sub>, 5-HT<sub>2C</sub>; muscarinic receptors M<sub>1</sub>, M<sub>2</sub>, M<sub>3</sub>, M<sub>4</sub>; adrenergic receptors beta1, beta2; dopamine receptors D<sub>2</sub>, D<sub>3</sub>, D<sub>4</sub>; and histamine receptor H<sub>1</sub>. All respective crystal structures were fetched from the PDB database [23], and they are presented in Table 1.

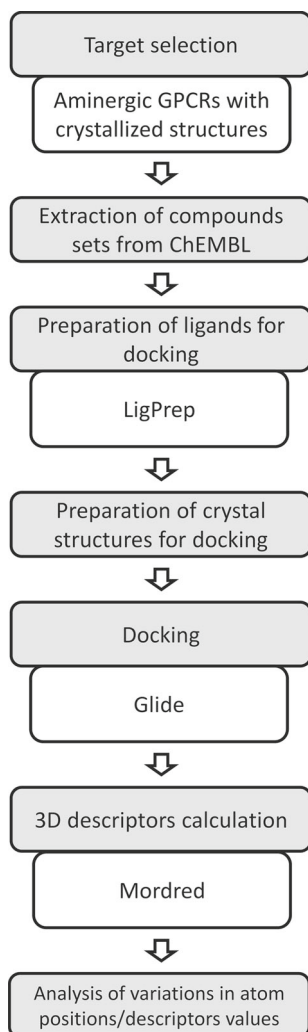
The compounds were prepared for docking in LigPrep [59] (protonation states generated at pH  $7.4 \pm 0.0$ ; a maximum of four stereoisomers per compound was allowed), and the number of compounds in each dataset is presented in Table 2.

The compounds were docked in Glide to the respective crystal structures; a maximum of ten docking poses were allowed for each compound. For each compound conformation obtained in docking, 3d descriptors were generated using the recently published package for descriptors calculation—Mordred [60]. It contains 214 3d descriptors grouped into the following categories: CPSA, geometrical index, gravitational index, MoRSE, moment of inertia.

The descriptor-based compound representations were compared in the following way—for each compound, the standard deviation (std) between values of all descriptors obtained for all docking poses of a particular compound was calculated and visualized using LargeVis by embedding into a 2d space. The results of such analysis were prepared in an interactive way and made available online, so chemical structures referring to a particular data point can be manually analyzed ([http://chem.gmunet.net/vischem\\_stability](http://chem.gmunet.net/vischem_stability)). In addition, studies considering the first  $k$  conformations (for compound sets sorted by docking score) were performed with  $k$  adopting values 3, 5, and 10. Moreover, the stds of coor-



**Fig. 1** Example analysis of consistency in compounds docked poses presented as the average std of all descriptors used for compounds representation



**Fig. 2** Scheme of the study presented in the paper

dinates of each atom of a compound were calculated (after the compounds alignment) and compared to determine how differences in descriptor values are related to differences in atom positions for various docking poses (Fig. 1).

The scheme of the whole project is presented in Fig. 2.

## Results and discussion

The averaged std values of generated descriptors between various compound conformations, depending on the number of compound orientations taken into account, are gathered in Table 3 with the stds of atom positions for docking poses obtained for a particular compound.

The analysis clearly shows that the variation of compound orientations in the binding site depends on the crystal structure used for docking rather than on the compound affinity for the receptor. Moreover, the rate of variation in the atom positions is not necessarily correlated with the rate of variation in the 3d descriptor values, although it is the most common situation. Additionally, active compounds were not more consistently docked (in terms of both 3d descriptor values and atom positions) in comparison with inactive molecules, although intuitively and theoretically that should be the case.

The number of docking poses taken into account in the performed analyses has only a slight impact on the observed dependencies that were rather consistent for a particular crystal structure, in terms of both similarity between 3d descriptor values and consistency of compound orientations described as the std of their atom coordinates. For the 5-HT<sub>1B</sub> crystal structures, the std of both 3d descriptor values and atom coordinates was higher for active than inactive compounds for all cases. On the other hand, 5-HT<sub>2B</sub> ligands were, in general, more consistent in terms of the analyzed properties than compounds that were inactive for this receptor.

Interestingly, 5-HT<sub>2C</sub> and ACM<sub>1</sub> ligands were more consistent compared with inactive compounds in terms of atom coordinates but less consistent when 3d descriptors were considered. However, for the rest of the muscarinic receptors, the tendencies were again shifted toward higher inconsistency for active compounds.

Adrenergic receptors were, in general, characterized by higher consistency for active compounds in terms of both parameters considered; however, there are examples of crystal structures for which there were variations in the consistency of 3d descriptor values and atom coordinates—such as 2R4R, 2R4S, 3KJ6, 3P0G, and 5D5A. The above-mentioned differences occurred also for the crystal structure of histamine receptor H<sub>1</sub>.

Example differences in docking poses for various crystal structures for two pairs of structurally similar active and inactive compounds (CHEMBL428892, CHEMBL66310, and CHEMBL387545, CHEMBL225364 pairs) toward 5-HT<sub>1B</sub> are presented in Fig. 3.

The docking results presented in Fig. 3 confirm the strong dependence of the consistency of a compound docking pose on the crystal structure used for docking. The active compound, CHEMBL428892, was docked in varying poses for the 4IAQ crystal structure, whereas when 4IAR and 5V54 were used, it similarly fit in the respective binding site.

**Table 3** Comparison of average std of between analyzed 3d descriptors and atom coordinates

Target	Crystal structure	3d descriptors std/std of atoms positions Number of conformations considered					
		3		5		10	
		Actives	Inactives	Actives	Inactives	Actives	Inactives
5-HT <sub>1B</sub>	4IAQ	<b>0.254/0.581</b>	0.208/0.536	<b>0.309/0.733</b>	0.252/0.650	<b>0.332/0.948</b>	0.271/0.846
	4IAR	<b>0.259/0.652</b>	0.212/0.571	<b>0.312/0.815</b>	0.260/0.692	<b>0.335/1.039</b>	0.278/0.880
	5V54	<b>0.246/0.582</b>	0.206/0.535	<b>0.295/0.719</b>	0.250/0.668	<b>0.320/0.890</b>	0.270/0.884
5-HT <sub>2B</sub>	4IB4	0.191/0.487	<b>0.219/0.539</b>	0.243/0.640	<b>0.266/0.675</b>	0.258/0.840	<b>0.281/0.853</b>
	4NC3	0.194/0.495	<b>0.217/0.526</b>	0.242/0.618	<b>0.272/0.671</b>	0.257/0.856	<b>0.290/0.816</b>
	5TUD	0.192/0.486	<b>0.217/0.524</b>	0.240/0.614	<b>0.266/0.656</b>	0.256/0.792	<b>0.282/0.807</b>
	5TVN	0.201/0.478	<b>0.225/0.529</b>	0.243/0.602	<b>0.272/0.657</b>	0.256/0.758	<b>0.285/0.838</b>
5-HT <sub>2C</sub>	6BQG	<b>0.204/0.503</b>	0.200/0.543	<b>0.246/0.627</b>	0.242/0.666	<b>0.258/0.755</b>	0.257/0.861
	6BQH	<b>0.211/0.518</b>	0.206/0.556	<b>0.258/0.650</b>	0.246/0.672	<b>0.272/0.804</b>	0.262/0.845
ACM <sub>1</sub>	5CXV	<b>0.255/0.572</b>	0.218/0.539	<b>0.304/0.716</b>	0.260/0.670	<b>0.325/0.907</b>	0.276/0.854
ACM <sub>2</sub>	3UON	<b>0.283/0.603</b>	0.207/0.534	<b>0.343/0.766</b>	0.246/0.663	<b>0.374/0.987</b>	0.260/0.813
	4MQS	<b>0.272/0.640</b>	0.197/0.506	<b>0.323/0.781</b>	0.244/0.643	<b>0.346/1.013</b>	0.255/0.845
	4MQT	<b>0.279/0.645</b>	0.202/0.509	<b>0.332/0.781</b>	0.242/0.643	<b>0.358/1.013</b>	0.257/0.804
ACM <sub>3</sub>	4DAJ	<b>0.284/0.587</b>	0.221/0.523	<b>0.340/0.736</b>	0.265/0.647	<b>0.359/0.998</b>	0.277/0.821
	4U14	<b>0.287/0.615</b>	0.220/0.521	<b>0.342/0.749</b>	0.262/0.646	<b>0.361/0.995</b>	0.276/0.824
	4U15	<b>0.288/0.598</b>	0.224/0.531	<b>0.343/0.747</b>	0.270/0.660	<b>0.362/0.982</b>	0.282/0.840
	4U16	<b>0.281/0.563</b>	0.220/0.533	<b>0.332/0.706</b>	0.268/0.671	<b>0.350/0.957</b>	0.281/0.844
ACM <sub>4</sub>	5DSG	<b>0.266/0.597</b>	0.195/0.506	<b>0.322/0.774</b>	0.240/0.626	<b>0.344/0.969</b>	0.252/0.797
Beta1	2VT4	0.246/0.555	<b>0.252/0.557</b>	0.290/0.675	<b>0.302/0.683</b>	0.304/0.813	<b>0.313/0.792</b>
	2Y00	0.237/0.528	<b>0.247/0.569</b>	0.283/0.691	<b>0.295/0.706</b>	0.294/0.813	<b>0.306/0.842</b>
	2Y01	0.252/0.624	<b>0.255/0.618</b>	0.299/0.755	<b>0.309/0.764</b>	0.312/0.884	<b>0.320/0.917</b>
	2Y02	<b>0.247/0.577</b>	0.246/0.570	0.293/0.704	<b>0.299/0.707</b>	0.306/0.806	<b>0.313/0.858</b>
	2Y03	0.252/0.570	<b>0.255/0.569</b>	0.299/0.698	<b>0.309/0.711</b>	0.311/0.878	<b>0.320/0.838</b>
	2YCW	0.240/0.514	<b>0.249/0.560</b>	0.285/0.652	<b>0.297/0.688</b>	0.296/0.773	<b>0.309/0.838</b>
	2YCX	0.246/0.547	<b>0.248/0.537</b>	<b>0.294/0.673</b>	0.291/0.633	<b>0.307/0.786</b>	0.305/0.830

Cases where std was higher for active compounds are presented in bold

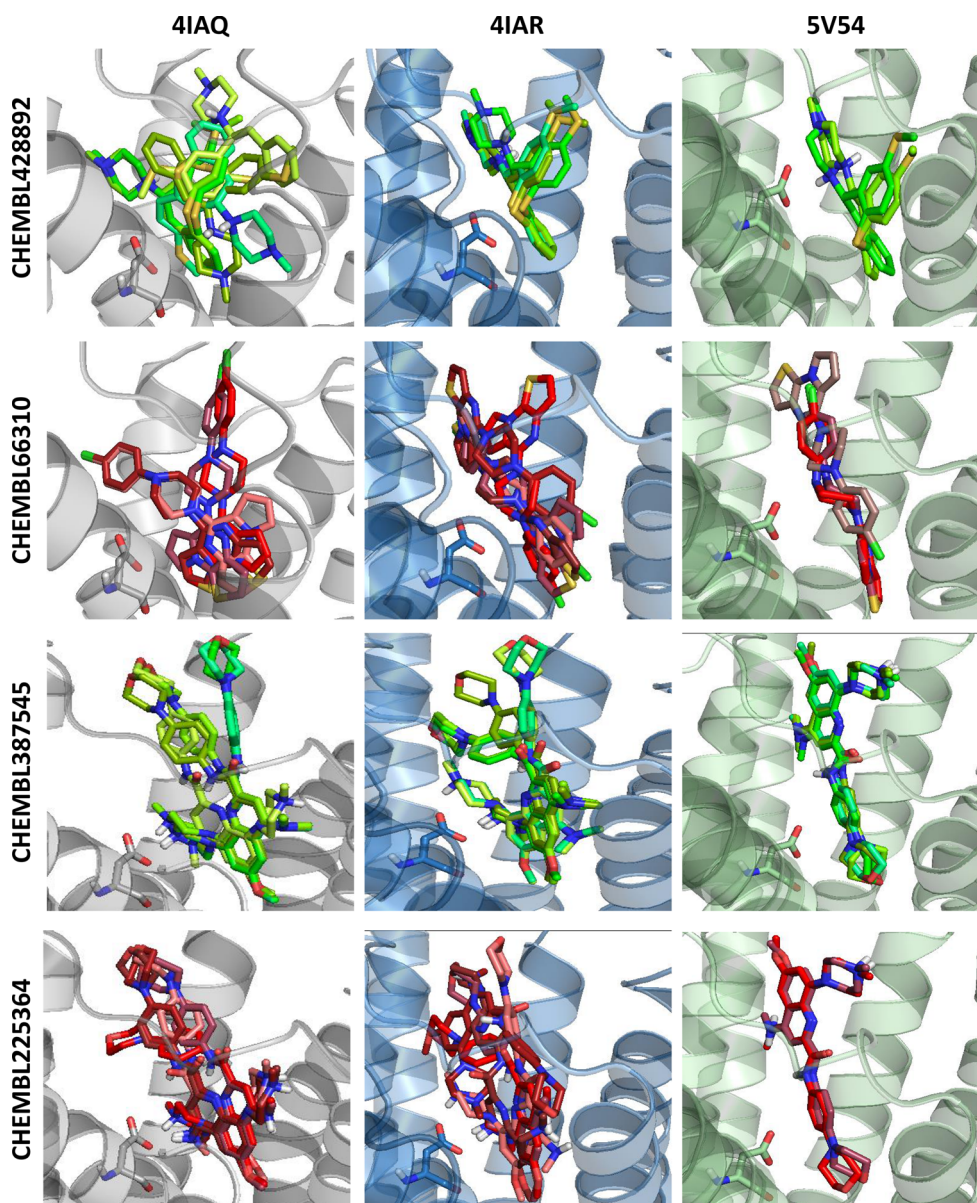
On the other hand, its inactive analog, ChEMBL66310, was docked less consistently to all crystals used; the highest variations occurred for 5V54, as one of the poses was flipped. ChEMBL387545 adopted two different orientations in the binding site of 4IAQ, and two in 4IAR, whereas for 5V54, all of the obtained docking poses were similar. ChEMBL225364, despite being inactive toward 5-HT<sub>1B</sub>, was docked very similarly to its 4IAQ and 5V54 crystal structures with conformation variations occurring for 4IAR.

Example visualization of the 3d descriptors space is presented in Fig. 4. It can be observed that for the outcome of this docking studies, representation by 3d descriptors will not necessarily provide good performance of predictive model constructed with the aim of making distinction between active and inactive compounds toward ACM<sub>3</sub>.

Visualizations for all crystal structures are available online at [http://chem.gmum.net/vischem\\_stability](http://chem.gmum.net/vischem_stability), together with

the provision of compound structures referring to a particular data point (Fig. 5). The possibility to manually analyze the results enables the explanation and optimization of machine learning methods performance and other automatic and semiautomatic analyses performed on 3d descriptors as compounds representation. Such visualizations can be of great help at the stage of designing such experiments, e.g., in terms of choosing the crystal structure that provides the best discrimination between active and inactive compounds, especially considering that the consistency in 3d descriptor space is not necessarily correlated with the std of atom coordinates in various docking poses, or in terms of proper evaluation of the applicability of a model, by the analysis of chemical structures of compounds that occupy similar regions of 3d descriptors space, despite possessing different activities toward considered receptor.





**Fig. 3** Analysis of docking results of selected compounds to crystal structures of serotonin receptor 5-HT<sub>1B</sub>; green: active compound, red: inactive compound; D3.32 residue of the protein is visualized as sticks. (Color figure online)

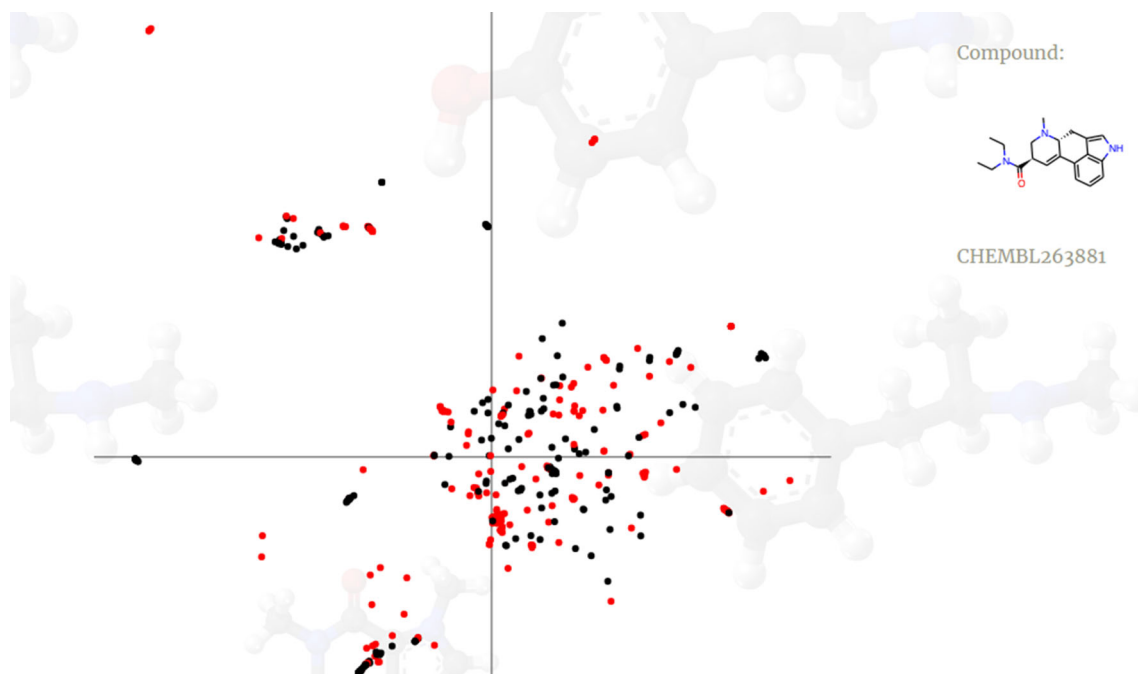
To explain the causes of such dependencies of the compound pose variations, the analysis of the relationship between the number of rotatable bonds and the std in 3d descriptor values, expressed as the Pearson correlation coefficient, was carried out (Table 4; for simplification, the results were averaged for all crystal structures for a particular protein). According to intuition, the performed analysis indicated a strong dependency of the consistency in descriptor values on the number of rotatable bonds (the higher the number of rotatable bonds, the lower the consistency of 3d descriptor values) with Pearson correlation values approaching 0.718 for 5-HT<sub>2B</sub>. Another observation is that in general, the correlation between the number of rotatable bonds and

the variations in compound orientations in the binding site (described by 3d descriptors) was stronger when a higher number of docked poses was taken into account. Example visualizations of such dependencies are presented in Fig. 6, and respective figures for the remaining data are present in the Supplementary Material.

Moreover, an analysis of std of each descriptor values separately was performed (Supplementary Material; for simplification, the obtained std values were averaged for all crystal structures available for a particular protein). It was found out that the lowest variation occurred for 3D-MoRSE descriptors determined for a distance equal to 1 (std of  $1-2 \times 10^{-18}$ ), whereas the highest std values were on aver-



**Fig. 4** Example analysis of stability of docking poses for ligands of the ACM<sub>3</sub> receptor. Gray area refers to compounds active or inactive toward other targets considered in the study. In this case, active compounds are relatively well demarcated from the inactive ones



**Fig. 5** Example visualization of the prepared online tool. Red points refer to active compounds and black to the inactive ones. Chemical structure corresponding to each data point can be manually analyzed. (Color figure online)

age obtained for descriptors from the group of geometrical indexes, such as the geometric Petitjean index and the geometrical shape index (std of 0.5–0.7 on average). Relatively high fluctuations in descriptor values (of similar std range as in the case of the geometrical shape index descriptors) were also observed for 3D-MoRSE descriptors calculated for a distance equal to 10.

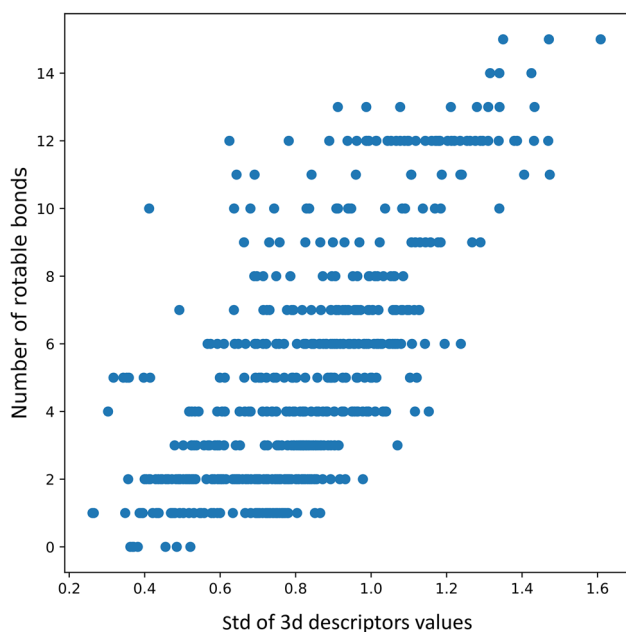
## Concluding remarks

In summary, a deep analysis of 3d descriptors generated for compounds docked to crystal structures of aminergic GPCRs, with the aim of their application in machine learning experiments, supplemented with the tool for manual inspection of structures referring to a particular data point was prepared. A strong dependence of the obtained results on the crystal



**Table 4** Comparison of values of the Pearson correlation coefficient between the number of rotatable bonds and the average std of the 3d descriptor values for particular compound conformations

Target	Number of compound conformations considered/Pearson correlation coefficient value		
	3	5	10
5-HT <sub>1B</sub>	0.367	0.452	0.546
5-HT <sub>2B</sub>	0.506	0.562	0.718
5-HT <sub>2C</sub>	0.442	0.519	0.618
ACM <sub>1</sub>	0.247	0.326	0.513
ACM <sub>2</sub>	0.289	0.316	0.562
ACM <sub>3</sub>	0.285	0.330	0.562
ACM <sub>4</sub>	0.265	0.462	0.657
Beta1	0.290	0.329	0.446
Beta2	0.185	0.211	0.332
D <sub>2</sub>	0.345	0.422	0.595
D <sub>3</sub>	0.283	0.334	0.530
D <sub>4</sub>	0.309	0.375	0.560
H <sub>1</sub>	0.190	0.218	0.000

**Fig. 6** Analysis of average std of the 3d descriptor values depending on the number of rotatable bonds in a particular compound for 5-HT<sub>2B</sub>R ligands, when ten docked poses were taken into account

structure used for docking was proved. Moreover, although the variation of 3d descriptor values is typically correlated with the variation of compound conformations, there were several cases where these dependencies were not preserved, revealing the limitations of the applied depiction of docking poses (for the ideal representation, higher variation in docking poses should lead to higher variation in descriptor values

and vice versa). Additionally, the ability to manually analyze all the information, including the analysis of chemical structures referring to particular data points, enables a better design of machine learning experiments conducted on such type of data, allowing for the maximization of the power of predictive models, for example by proper selection of crystal structures for studies (those that provide the best discrimination between active and inactive compounds in machine learning-based experiments) or the proper evaluation of the model applicability via the analysis of chemical structures with overlapping fragments of descriptors space.

**Acknowledgements** The study was supported by the National Science Centre, Poland, within the HARMONIA 7 Grant 2015/18/M/NZ7/00377.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN (2007) Virtual screening in drug discovery—a computational perspective. *Curr Prot Pept Sci* 4:329–351. <https://doi.org/10.2174/138920307781369427>
- Breda A, Basso LA, Santos DS, de Azevedo WF Jr. (2008) Virtual screening of drugs: score functions, docking, and drug design. *Curr Comput Aided Drug Des* 4:265–272. <https://doi.org/10.2174/157340908786786047>
- Nicholls A (2008) What do we know and when do we know it? *J Comput Aided Mol Des* 3–4:239–255. <https://doi.org/10.1007/2Fs10822-008-9170-2>
- Geppert H, Vogt M, Bajorath J (2010) Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J Chem Inf Model* 2:205–216. <https://doi.org/10.1021/ci900419k>
- Gao Q, Yang L, Zhu Y (2010) Pharmacophore based drug design approach as a practical process in drug discovery. *Curr Comput Aided Drug Des* 1:37–49. <https://doi.org/10.2174/157340910790980151>
- Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 11–12:444–450. <https://doi.org/10.1016/j.drudis.2010.03.013>
- Anderson AC (2003) The process of structure-based drug design. *Chem Biol* 9:787–797. <https://doi.org/10.1016/j.chembiol.2003.09.002>
- Heikamp K, Bajorath J (2011) How do 2D fingerprints detect structurally diverse active compounds? Revealing compound subset-specific fingerprint features through systematic selection. *J Chem Inf Model* 9:2254–2265. <https://doi.org/10.1021/ci200275m>
- Duan J, Sastry M, Dixon S, Lowrie J, Sherman W (2010) Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J Mol Gr Model* 2:157–170. <https://doi.org/10.1016/j.jmglm.2010.05.008>
- Clark RD, Patterson DE, Soltanshahi F, Blake JF, Matthew JB (2000) Visualizing substructural fingerprints. *J Mol Gr*

- Model 4–5(404–411):527–532. [https://doi.org/10.1016/S1093-3263\(00\)00065-6](https://doi.org/10.1016/S1093-3263(00)00065-6)
11. Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2004) Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org Biomol Chem* 22:3256–3266. <https://doi.org/10.1039/B409865J>
  12. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* 1:195–207. <https://doi.org/10.1021/ci600342e>
  13. Deng Z, Chuaqui C, Singh J (2006) Knowledge-based design of target-focused libraries using protein–ligand interaction constraints. *J Med Chem* 2:490–500. <https://doi.org/10.1021/jm050381x>
  14. Bender A, Jenkins JL, Scheiber J, Sukuru SCK, Glick M, Davies J (2009) W. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J Chem Inf Model* 1:108–119. <https://doi.org/10.1021/ci800249s>
  15. Willett P, Barnard JM, Downs GM, Information BC, Road U, Sheffield S (1998) Chemical similarity searching. *J Chem Inf Comput Sci* 6:983–996. <https://doi.org/10.1021/ci9800211>
  16. Acharya C, Coop A, Polli JE, MacKerell AD Jr (2011) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des* 7:10–22. <https://doi.org/10.2174/157340911793743547>
  17. Lounnas V, Ritschel J, Kelder J, McGuire R, Bywater RP, Foloppe N (2013) Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* 5:e201302011. <https://doi.org/10.5936/csbj.201302011>
  18. Mitchell JBO (2014) Machine learning methods in chemoinformatics. *Wiley Interdiscip Rev Comput Mol Sci* 5:468–481. <https://doi.org/10.1002/2fwcms.1183>
  19. Melville JL, Burke EK, Hirst JD (2009) Machine learning in virtual screening. *Comb Chem High Throughput Screen* 4:332–343. <https://doi.org/10.2174/138620709788167980>
  20. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2011) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40:D1100–D1107. <https://doi.org/10.1093/nar/gkr777>
  21. Warszycki D, Mordalski S, Kristiansen K, Kafel R, Sylte I, Chilmoneczek Z, Bojarski AJ (2013) A linear combination of pharmacophore hypotheses as a new tool in search of new active compounds—an application for 5-HT<sub>1A</sub> receptor ligands. *PLoS ONE* 8:e84510. <https://doi.org/10.1371/journal.pone.0084510>
  22. Kalliokoski T, Kramer C, Vulpetti A, Gedeck P (2013) Comparability of mixed IC<sub>50</sub> data—a statistical analysis. *PLoS ONE* 8:e61007. <https://doi.org/10.1371/journal.pone.0061007>
  23. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
  24. Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V, Han GW, Liu W, Huang X-P, Vardy E, McCorvy JD, Gao X, Zhou XE, Melcher K, Zhang C, Bai F, Yang H, Yang L, Jiang H, Roth BL, Cherezov V, Stevens RC, Xu HE (2013) Structural basis for molecular recognition at serotonin receptors. *Science* 340:610–614. <https://doi.org/10.1126/science.1232807>
  25. Yin WC, Zhou XE, Yang D, De Waal P, Wang MT, Dai A, Cai X, Huang CY, Liu P, Yin Y, Liu B, Caffrey M, Melcher K, Xu Y, Wang MW, Xu HE, Jiang Y (2018) Crystal structure of the human 5-HT<sub>1B</sub> serotonin receptor bound to an inverse agonist. *Cell Discov* 4:12. <https://doi.org/10.1038/s41421-018-0009-2>
  26. Wacker D, Wang C, Katritch V, Han GW, Huang X-P, Vardy E, McCorvy JD, Jiang Y, Chu M, Siu FY, Liu W, Xu HE, Cherezov V, Roth BL, Stevens RC (2013) Structural features for functional selectivity at serotonin receptors. *Science* 340:615–619. <https://doi.org/10.1126/science.1232808>
  27. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, Zatssepin NA, Li D, Messerschmidt M, Boutet S, Williams GJ, Koglin JE, Seibert MM, Wang C, Shah STA, Basu S, Fromme R, Kupitz C, Rendek KN, Grotjohann I, Fromme P, Kirian RA, Beyerlein KR, White TA, Chapman HN, Caffrey M, Spence JCH, Stevens RC, Cherezov V (2013) Serial femtosecond crystallography of G protein-coupled receptors. *Science* 342:1521–1524. <https://doi.org/10.1126/science.1244142>
  28. Ishchenko A, Wacker D, Kapoor M, Zhang A, Han GW, Basu S, Patel N, Messerschmidt M, Weierstall U, Liu W, Katritch V, Roth BL, Stevens RC, Cherezov V (2017) Structural insights into the extracellular recognition of the human serotonin 2B receptor by an antibody. *Proc Natl Acad Sci* 114:8223–8228. <https://doi.org/10.1073/pnas.1700891114>
  29. Wacker D, Wang S, McCorvy JD, Betz RM, Venkatakrishnan AJ, Levit A, Lansu K, Schools ZL, Che T, Nichols DE, Shoichet BK, Dror RO, Roth BL (2017) Crystal structure of an LSD-bound human serotonin receptor. *Cell* 168:377–389. <https://doi.org/10.1016/j.cell.2016.12.033>
  30. Peng Y, McCorvy JD, Harpsøe K, Lansu K, Yuan S, Popov P, Qu L, Pu M, Che T, Nikolajsen LF, Huang XP, Wu Y, Shen L, Bjørn-Yoshimoto WE, Ding K, Wacker D, Han GW, Cheng J, Katritch V, Jensen AA, Hanson MA, Zhao S, Gloriam DE, Roth BL, Stevens RC, Liu ZJ (2018) 5-HT<sub>2C</sub> receptor structures reveal the structural basis of GPCR polypharmacology. *Cell* 172:719–730. <https://doi.org/10.1016/j.cell.2018.01.001>
  31. Thal DM, Sun B, Feng D, Nawaratne V, Leach K, Felder CC, Bures MG, Evans DA, Weis WI, Bachhawat P, Kobilka TS, Sexton PM, Kobilka BK, Christopoulos A (2016) Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. *Nature* 531:335–340. <https://doi.org/10.1038/nature17188>
  32. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482:547–551. <https://doi.org/10.1038/nature10753>
  33. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hübner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504:101–106. <https://doi.org/10.1038/nature12735>
  34. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482:552–556. <https://doi.org/10.1038/nature10867>
  35. Thorsen TS, Matt R, Weis WI, Kobilka BK (2014) Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis. *Structure* 22:1657–1664. <https://doi.org/10.1016/j.str.2014.08.022>
  36. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX (2008) Structure of a  $\beta$ 1-adrenergic G-protein-coupled receptor. *Nature* 454:486–491. <https://doi.org/10.1038/nature07101>
  37. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG, Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a  $\beta$ 1-adrenergic receptor. *Proc Natl Acad Sci* 108:8228–8232. <https://doi.org/10.1073/pnas.1100185108>
  38. Christopher JA, Brown J, Doré AS, Errey JC, Koglin M, Marshall FH, Myszkowski DG, Rich RL, Tate CG, Tehan B, Warne T, Congreve

- M (2013) Biophysical fragment screening of the  $\beta$ 1-adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design. *J Med Chem* 56:3446–3455. <https://doi.org/10.1021/jm400140q>
39. Warne T, Edwards PC, Leslie AGW, Tate CG (2012) Crystal structures of a stabilized  $\beta$ 1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* 20:841–849. <https://doi.org/10.1016/j.str.2012.03.014>
  40. Miller-Gallacher JL, Nehmé R, Warne T, Edwards PC, Schertler GFX, Leslie AGW, Tate CG (2014) The 2.1 Å resolution structure of cyanopindolol-bound  $\beta$ 1-adrenoceptor identifies an intramembrane Na<sup>+</sup> ion that stabilises the ligand-free receptor. *PloS ONE* 9:e92727. <https://doi.org/10.1371/journal.pone.0092727>
  41. Huang J, Chen S, Zhang JJ, Huang XY (2013) Crystal structure of oligomeric  $\beta$ 1-adrenergic G protein-coupled receptors in ligand-free basal state. *Nat Struct Mol Biol* 20:419–425. <https://doi.org/10.1038/nsmb.2504>
  42. Sato T, Baker J, Warne T, Brown GA, Leslie AGW, Congreve M, Tate CG (2015) Pharmacological analysis and structure determination of 7-methylcyanopindolol-bound  $\beta$ 1-adrenergic receptor. *Mol Pharmacol* 88:1024–1034. <https://doi.org/10.1124/mol.115.101030>
  43. Leslie AGW, Warne T, Tate CG (2015) Ligand occupancy in crystal structure of  $\beta$ 1-adrenergic G protein-coupled receptor. *Nat Struct Mol Biol* 22:941–942. <https://doi.org/10.1038/nsmb.3130>
  44. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450:383–387. <https://doi.org/10.1038/nature06325>
  45. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318:1258–1265. <https://doi.org/10.1126/science.1150577>
  46. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EYT, Velasquez J, Kuhn P, Stevens RC (2008) A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure* 16:897–905. <https://doi.org/10.1016/2Fj.str.2008.05.001>
  47. Bokoch MP, Zou Y, Rasmussen SGF, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi HJ, Thian FS, Kobilka TS, Puglisi JD, Weis WI, Pardo L, Prosser RS, Mueller L, Kobilka BK (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463:108–112. <https://doi.org/10.1038/nature08650>
  48. Wacker D, Fenalti G, Brown MA, Katritch V, Abagyan R, Cherezov V, Stevens RC (2010) Conserved binding mode of human  $\beta$ 2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J Am Chem Soc* 132:11443–11445. <https://doi.org/10.1021/ja105108q>
  49. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konezki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the  $\beta$ (2) adrenoceptor. *Nature* 469:175–180. <https://doi.org/10.1038/nature09648>
  50. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SGF, Choi H-J, DeVree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK (2011) Structure and function of an irreversible agonist- $\beta$ (2) adrenoceptor complex. *Nature* 469:236–240. <https://doi.org/10.1038/2Fnature09665>
  51. Zou Y, Weis WI, Kobilka BK (2012) N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. *PLoS ONE* 7:e46039. <https://doi.org/10.1371/journal.pone.0046039>
  52. Huang CY, Olieric V, Ma P, Howe N, Vogeley L, Liu X, Warshamange R, Weinert T, Panepucci E, Kobilka B, Diederichs K, Wang M, Caffrey M (2016) In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. *Struct Biol* 72:93–112. <https://doi.org/10.1107/S2059798315021683>
  53. Ma P, Weichert D, Aleksandrov LA, Jensen TJ, Riordan JR, Liu X, Kobilka BK, Caffrey M (2017) The cubicon method for concentrating membrane proteins in the cubic mesophase. *Nat Protoc* 12:1745–1762. <https://doi.org/10.1038/nprot.2017.057>
  54. Liu X, Ahn S, Kahsai AW, Meng KC, Latorraca NR, Pani B, Venkatakrishnan AJ, Masoudi A, Weis WI, Dror RO, Chen X, Lefkowitz RJ, Kobilka BK (2017) Mechanism of intracellular allosteric  $\beta$ 2AR antagonist revealed by X-ray crystal structure. *Nature* 548:480–484. <https://doi.org/10.1038/nature23652>
  55. Wang S, Che T, Levit A, Shoichet BK, Wacker D, Roth BL (2018) Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* 555:269–273. <https://doi.org/10.1038/nature25758>
  56. Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330:1091–1095. <https://doi.org/10.1126/science.1197410>
  57. Wang S, Wacker D, Levit A, Che T, Betz RM, McCorvy JD, Venkatakrishnan AJ, Huang XP, Dror RO, Shoichet BK, Roth BL (2017) D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* 358:381–386. <https://doi.org/10.1126/science.aan5468>
  58. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65–70. <https://doi.org/10.1038/nature10236>
  59. LigPrep, Schrodinger Release 2018-1, LLC, New York, NY, 2018
  60. Moriwaki H, Tian Y-S, Kawashita N, Takagi T (2018) Mordred: a molecular descriptor calculator. *J Cheminf* 10:4. <https://doi.org/10.1186/s13321-018-0258-y>